RESPONSE LINKED CENSORING: MODELING AND ESTIMATION

BY

DONALD R. HOOVER and FRANK M. GUESS

TECHNICAL REPORT NO. 422
SEPTEMBER 14, 1989

PREPARED UNDER CONTRACT
N00014-89-J-1627 (NR-042-267)
FOR THE OFFICE OF NAVAL RESEARCH

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

4

RESPONSE LINKED CENSORING: MODELING AND ESTIMATION

BY

DONALD R. HOOVER and FRANK M. GUESS

TECHNICAL REPORT NO. 422
SEPTEMBER 14, 1989

Prepared Under Contract
N00014-89-J-1627 (NR-042-267)
For the Office of Naval Research

Herbert Solomon, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

1. INTRODUCTION

Survival data (or more broadly, time to response data) often has censored values. If censoring is independent of the response, then methods to analyze the data are well known and documented. The product limit (Kaplan Meier) estimator can be used for a nonparametric approach; the Cox proportional hazard model yields what is sometimes called a semiparametric analysis and, of course, parametric techniques also exist. Cf. for example: Cox and Oakes (1984), Lawless (1982), and Kalbfleisch and Prentice (1980).

Slud and Byar (1988) point out that for many biomedical experiments the assumption of independence is reasonable. This is especially true in clinical trials where the mechanism of loss to follow-up is end of study censoring and there is no systematic pattern of patient entry into the study over time.

For other situations, Slud and Byar (1988) emphasize that the assumption of independence may not be plausible and for many commonly occurring problems, even independence conditional on known covariates does not hold. It is, therefore, imperative to develop methods of modeling and analyzing such situations.

One impediment to development of such approaches is the nonidentifiability of dependence between censoring and response (see Tsiatis (1975)). That is, without making some assumptions, it is impossible to determine from the data alone whether or not the censoring time and response time are independent.

Klein and Moeschberger (1988) used a particular dependence structure to obtain bounds on the marginal survival. Slud and Rubinstein (1983) and Peterson (1976) have developed other bounds under dependent censoring. See Moeschberger and Klein (1984), Basu and Klein (1982) and David and Moeschberger (1978) for more on dependent censoring (competing risks).

A particular type of dependent censoring defined as response linked (RL) censoring will now be introduced and modeled. RL censoring is censoring caused

by occurrence of the response or the fact that the response is about to occur. For instance, consider a heroin addiction rehabilitation study where the response is a relapse to heroin addiction. It may be the case that some subjects restarting heroin usage also stop going to the clinic for chemical testing and hence become censored. RL censoring may also be generated by other unknown (or unmeasurable) factors which cause both the response and censoring to be linked. E.g., such a factor could be an ineffective counselor at the clinic causing the subject to both leave the study and relapse to addiction.

As another example, consider a wildlife study on animals that have been radio tagged. E.g., see Pollock et al (1988a), and Pollock, Winterstein, and Conroy (1988b). A particular bird (duck or bobwhite) is followed until the time it is known to die (the radio signal remains in one location) or is censored (the radio is destroyed or the signal is unreceivable). The bird could be destroyed along with the radio due to a predator or some other cause. This would be a RL censoring with the unobserved death occurring at or near the time of censoring. Another animal's radio receiver might simply malfunction (from a manufacturing flaw) to result in what would be an independent censoring.

RL censoring occurs (possibly together with independent censoring) in a variety of settings. For example, programs to [illegible] smoking (cocaine, crack, etc); studies of methods to reduce the time between violent behaviors for high security prisoners (e.g., furlough versus lockdown); clinical trials on humans (e.g., use of prescribed contraceptives); etc.

Sometimes a censoring will be known to be independent, e.g., the end of study censoring mentioned earlier. Such cases of known independent censorings will be referred to by (E). Often it will be impossible to distinguish between RL and independent censoring. Censorings which cannot be classified as RL or independent will be referred to as unknown censoring (U). These (U) censorings

will include both (RL) and nondistinguishable independent (NI) censorings. Responses will be noted by (R).

In the next section, a parametric model will be developed that incorporates these dual components of the unknown censorings to develop parametric estimates.

2. THE RESPONSE LINKED CENSORING MODEL

For notational purposes, let $j = 1,\ldots, n$ denote observations while:

$\underset{\sim}{Z}_j = (Z_{j1}, Z_{j2},\ldots, Z_{jk})$ is a vector of information from k covariates.

R_j is the time of response. The probability density function (pdf) and cumulative density function (cdf) of time to R_j, may depend on the covariates and are denoted to by $f(t_j|\underset{\sim}{Z}_j)$ and $F(t_j|\underset{\sim}{Z}_j)$ respectively.

η_j is 1 if the response would be observed at the R_j and 0 otherwise. For any observation $P\{\eta_j = 1\}$ is P and is independent of $\underset{\sim}{Z}_j$.

N_j is the time of NI censoring. The time to N_j is independent of the covariates. The pdf and cdf of this time are referred to by $h(t_j)$ and $H(t_j)$ respectively.

E_j is the time of E censoring. It does not matter if the time to E_j is dependent on the covariates. The pdf and cdf of this time are referred to by $g(t_j)$ and $G(t_j)$ respectively.

R_j, η_j, E_j and $N_j|\underset{\sim}{Z}_j$ are independent of each other and of R_i, η_i, E_i and $N_i|\underset{\sim}{Z}_i$ for any i, j.

For simplicity, it has been assumed that any RL censoring is observed at the time of death. While this won't always be true (i.e., a wounded animal could survive 2 weeks after its radio was destroyed in an attack), any error this assumption introduces will likely be small compared to the error inherent in identifying the exact time of response. See Section 5 for suggestions about allowing for a more general RL censoring process. It also has been assumed that the probability of an otherwise uncensored response being RL censored is (1-P) regardless of the time of response or value of covariates. To assume otherwise

would vastly complicate the model.

[B The investigator will not observe $(R_j, \eta_j, N_j, E_j, \underset{\sim}{Z}_j)$ but rather will only see $(T_j, C_j, \underset{\sim}{Z}_j)$ where

T_j is the time of occurrence for first event which is min (R_j, E_j, N_j).

C_j is the censor status of the first event which by convention is

3 if that event is an E censoring

2 if that event is a U censoring (i.e., either the first event is R and η_j is 0 or the first event is NI)

1 if that event is a non RL censored response (i.e., η_j must be 1 and the first event is R).

Note that in the event of ties, an E censoring (usually caused by end of study) would be observed before a U censoring and any type of censoring would be observed before a response.

In terms of f, F, g, G, h, H, P and $\underset{\sim}{Z}_j$ the likelihood function for all observed value of (T_j, C_j) is:

$$(*) \quad \prod_{R \text{ observed}} [P \cdot f(t_j|\underset{\sim}{Z}_j)\,(1 - H(t_j))\,(1 - G(t_j))] \times$$

$$\prod_{U \text{ observed}} [(1-P) \cdot f(t_j|\underset{\sim}{Z}_j)\,(1-H(t_j))\,(1-G(t_j)) + h(t_j)\,(1-F(t_j|\underset{\sim}{Z}_j))\,(1-G(t_j))] \times$$

$$\prod_{E \text{ observed}} [g(t_j)\,(1-F(t_j|\underset{\sim}{Z}_j)\,(1-H(t_j))]$$

The U censorings consist of two components (RL censorings and NI censorings) that without any assumptions are indistinguishable from each other. The assumption of independence between NI censorings and the covariates makes it somewhat possible to separate these components. This assumption means that the relation between time and covariates for the NI component of observed U censorings will be opposite of the relationship between covariates and time in observed deaths. This is because the NI component will not be observed if the response occurs first. The relationship between covariates and time for the RL

component of observed U censorings should be the same as the relationship between response and time for observed responses. In Section 3, the likelihood representation in (*) will be used to estimate the contribution of RL components to the U censoring.

The importance of the assumption of independence between covariates and NI censoring cannot be overlooked because it implies that any relationship between the time of U censoring and the covariates which is in the same direction as the relationship between the time of response and the covariates is being caused by RL censoring. Sometimes this assumption may be appropriate and other times not. For instance, consider a cancer survival study where outcome is death by cancer and the U censoring is death from other causes (which may or may not be attributable to the cancer.) If the covariate is age, then it is natural that older individuals are more likely to die sooner of cancer and of other causes (independently of cancer) than are younger individuals. If, however, the covariate is severity of initial diagnosis of cancer, it seems reasonable that any association between this and early time of "other causes" death is due to either misdiagnosis or the cancer itself otherwise contributing to the death.

Therefore, an important strategy in RL analysis would be selection of covariates that have little or no logical association with NI censoring. If one is uncertain of association of covariates with NI censoring then they should analyze the data both allowing and not allowing for RL censoring and compare the results of each analysis. In Section 3 it is shown how to do such analysis in a parametric setting.

3. ANALYSIS OF RL CENSORING IN A PARAMETRIC SETTING

Parametric modeling of response time has often proven helpful in medicine and biology. For example, the Gompertz and Makeham, distributions have a long rich history (going back to the 1800's) in modeling lifelengths. See Cox and

Oakes (1984) and Gross and Clarke (1975) with their related references to these and other parametric models such as the Weibull and exponential. These authors also give advice on picking the appropriate parametric model. Miller (1983) and Efron (1988) point out that a parametric approach to modeling survival time is often more efficient compared to nonparametric approaches. Competing risks (censored data) has been modeled parametrically. Cf. Moeschberger and Klein (1984), Basu and Klein (1982), and David and Moeschberger (1978).

Let $\underset{\sim}{\alpha}$ be the vector of parameters associated with F and $\underset{\sim}{\beta}$ be the vector of parameters associated with H. Then the log likelihood function is

$$l = \sum_{R \text{ observed}} \ln(P) + \ln(f(t_j|\underset{\sim}{Z}_j, \underset{\sim}{\alpha}) + \ln(1-H(t_j|\underset{\sim}{\beta}) + \ln(1-G(t))$$

$$+ \sum_{U \text{ observed}} \ln[(1-P)\, f(t_j|\underset{\sim}{Z}_j, \underset{\sim}{\alpha})\, (1-H(t_j|\underset{\sim}{\beta})) +$$

$$h(t_j|\underset{\sim}{\beta})\, (1-F(t_j|\underset{\sim}{Z}_j, \underset{\sim}{\alpha}))] + \ln(1-G(t_j))$$

$$+ \sum_{E \text{ observed}} \ln(g(t_j)) + \ln(1-F(t_j|\underset{\sim}{Z}_j, \underset{\sim}{\alpha})) + \ln(1-H(t,|\underset{\sim}{\beta}))$$

The parameters of l are P, $\underset{\sim}{\alpha}$ and $\underset{\sim}{\beta}$. Provided that the likelihood function satisfies the assumptions on page 409 of Lehman (1983) then standard likelihood analysis, methods such as those in Miller (1982), may be used to derive estimates and test hypothesis of the parameters. Note that g and G are independent of the parameters and hence will drop out of any maximization of l.

An important hypothesis pair concerns the existence of RL censoring. The null hypothesis of no RL censoring is H_0:P=1 and the alternative of some RL censoring is the one sided H_a:P < 1. Let l_w be the maximum likelihood with P = 1 and l_Ω be the maximum likelihood with P unrestricted and $\hat{p}$ be the value of P producing this unrestricted maximum likelihood. It follows from the asymptotic theory in Miller (1981), that if H_0 is true, then

$2[l_\Omega - l_w]$ will have an approximate χ^2_1 distribution and that

$(-1)^{I(\hat{p} > 1)} [2(l_Q - l_w)]^{\frac{1}{2}} \sim N(0,1)$ where I is the standard indicator function of the set. Thus a test of H_0 vs. H_A with approximate size α is given by:

$$\begin{cases} \text{Reject } H_0 \text{ if } \hat{p} < 1 \text{ and } \sqrt{2(l_Q - l_w)} > Z_{(1-\alpha)} \\ \text{Do not reject } H_0 \text{ otherwise} \end{cases}$$

In Section 4 hypothesis testing of RL censoring and parameter estimates in the presence and absence of RL censoring will be illustrated.

4. EXAMPLES

Models using RL censoring are applied to three examples from the literature; on the Stanford heart transplant data, on rehabilitated duck survival data, and on lung cancer data. References and details are given below. In each of these examples, the U censoring time seems to be consistent with an exponential model. For simplicity, $\underset{\sim}{Z}$ will here always consist of only one covariate thus be denoted Z. Note that $F(t_j|Z,\underset{\sim}{\alpha})$ and $H(t_j|\underset{\sim}{\beta})$ will have the following forms:

$$F(t_j|Z,\underset{\sim}{\alpha}) = 1-e^{-t_j e^{[\alpha_0+\alpha_1 Z_j]}} \quad \text{for } -\infty < \alpha_0, \alpha_1 < \infty$$

$$H(t_j|\underset{\sim}{\beta}) = 1-e^{-t_j e^{\beta_0}} \quad \text{for } -\infty < \beta_0 < \infty$$

Compare Kalbfleisch and Prentice (1980) for similar forms.

It will be appropriate to use standard maximum likelihood analysis (see Miller (1981) and Lehman (1983)) with the above F and H in the parametric RL model.

Example One: Stanford Heart Transplant Data

The Stanford heart transplant data set in Miller (1981) has 69 patients. The patient deaths were classified into two groups: Heart rejection (considered

the response) and nonrejection (considered U censoring here). Many of the patients were alive (E censoring) at the time the data set was compiled. Except for the first two weeks following transplant, which had an abundance of nonrejection deaths, the times of response and U censoring appear consistent with an exponential model. The covariate was mismatch score between heart donor and recipient tissues. Miller (1981) examined several approaches each of which found higher mismatch scores to be associated with earlier rejection deaths. One would not expect mismatch scores to be associated with nonrejection deaths (U censoring) unless either the death types were misclassified or heart rejection contributed to the nonrejection deaths.

Elimination of individuals that died or were censored within two weeks of transplant and/or did not have mismatch scores reduced the sample to 57 individuals (28 of which died from rejection deaths and 7 of whom died from nonrejection deaths). Exponential models allowing for RL censoring and excluding RL censoring (P constrained to 1) were fitted to survival time in fraction of years beyond 14 days. Table I gives the results.

Table -

RL Censoring and Stanford Heart Transplant Data

Parameter Estimates	with RL Censoring	without RL Censoring
$\dot{\alpha}_0$	2.42	2.27
$\dot{\alpha}_1$	-1.41	-1.22
$\dot{\beta}_0$	2.49	2.09
$\dot{P}$	0.085	1
log likelihood (1)	70.54	71.55

Using the likelihood ratio test for $H_0: P=1$ vs. $H_a: P < 1$ would reject H_0 for the α error less than or equal to 0.0775. Including RL censoring lowered the estimate of the covariate coefficient ($\hat{\alpha}_1$) from -1.41 to -1.22 thus increasing the magnitude of the effect from higher mismatch scores on earlier deaths. It should be mentioned that after incorporation of more patients and follow up in the heart transplant data, Miller and Halperin (1982) decided to combine together the rejection and nonrejection deaths due to difficulties and arbitrariness in making the distinction between them.

Example Two: Rehabilitated Duck Survival

Fifty radio tagged released female black ducks were followed until censoring or natural death (Pollock et al (1988b)). Of these birds, 8 died naturally, 10 were shot by hunters (U censored) and the rest were end of study censored. Exponential models are used to model released animal survival time, e.g., Trent and Rongstad (1974). Pollock et al assumed random censoring in showing that a covariate, condition index, was statistically related to survival time.

It might be possible that some of the 10 birds shot were already in bad shape and would have died from natural causes soon. (The poor condition making the birds easier to be hunted.) This idea was tested by fitting a mixed exponential likelihood model to the data. The MLE for $\hat{p}$ was negative. This implies that any relationship between condition index and hunter censoring was opposite to the relationship between condition index and survival; and thus, shooting deaths are not RL censoring.

Example Three: Lung Cancer Survival

Prentice (1973) fits an exponential model to survival times of 40 lung patients (37 died during the study, 3 were end of study censored and there was

no U censoring). A covariate called performance status was strongly related to survival time.

Here, three simulation studies to test the efficacy of the mixed censoring model on data where the U censoring is known to be entirely RL were done with this data. The simulations involved fitting mixed censoring and all independent censoring models to 50 replications of each of the following alterations: (i) 5 randomly chosen responses changed to U censorings. (ii) 9 randomly chosen responses changed to U censorings. (iii) 18 randomly chosen responses changed to U censorings. (This simulation approach is analogous to Zippin and Armitage (1966).) Note that for the above data sets created in (i), (ii) and (iii), all of the U censorings will actually be RL censorings.

Table II summarizes the ability of the mixed model to detect the RL censoring in the simulation studies. In all three studies, $\hat{p}$ was positive in over 80% of the replications. The likelihood ratio test for $\hat{p}=0$ was usually statistically significant at $\alpha = 0.05$ in case (i), at $\alpha = 0.01$ in case (ii), and at $\alpha = 0.001$ in case (iii). This is in spite of the small sample size of 40 observations.

Table III examines to what degree incorporation of RL censoring in the mixed model improves the estimation of coefficient of the covariate performance status. The coefficient of this covariate in the exponential model fitted to the unaltered lung cancer data set was $\hat{B}_1 = 0.0600$. The coefficients in the simulation models are examined in terms of how close they are to 0.0600 (closer means superior).

The all independent censoring model seldom (always less than 20%) produced superior estimates $\hat{B}_1$ than did the mixed model for any of the replications in (i), (ii) and (iii). The mixed model gave superior estimates of $\hat{B}_1$ in over 60% of the replicates in each of the simulation studies. In case (iii) the mixed model estimate of $\hat{B}_1$ was superior to the independent model estimate in 98% of

the replicates. The mixed model estimated $\hat{B}_1$ to be 0.0600 (the value of $\hat{B}_1$ for unaltered data set) in 60% to 80% of the replicates, compared to never for the independent censoring model. The mixed model $\hat{B}_1$'s had mean absolute deviations from 0.0600 which were 40% to 80% smaller than those of the corresponding independent censoring models.

The mean absolute deviations from 0.0600 of the coefficients estimated by the mixed model stayed between .0015 and .0020 and did not appear to be related to the number of observations randomly censored. On the other hand, the mean absolute deviations from 0.0600 of the coefficients estimated by the independent models were almost directly proportional to the number of observations randomly censored growing from .00338 for 5 random censorings to .00932 for 18 random censorings. Still with 18/37 or close to 50% of the non RL censored observations randomly censored, the ratio of the mean absolute deviation to the unaltered sample coefficient value was only .00932/.0600 = .20.

Table II

Mixed Model vs. Independent Model

Detecting the RL Censoring

[Likelihood ratio test of H_0: p = 0 values]

Number of Responses Randomly Censored	Percent of LR P-values < .001	Percent of LR P-values < .01	Percent of LR P-values < .05	Percent of LR P-value < .50 (same as $\hat{p} > 0$)
5	10%	38%	66%	82%
9	18%	58%	76%	92%
18	82%	96%	100%	100%

Table III

Comparing the Mixed and Independent Model

Estimates for Coefficient of the Covariate

Performance Status

Number of Responses Randomly Censored	Percent Independent Model Estimate was closer to Unaltered Data Value	Percent Mixed Model Estimate was closer to Unaltered Data Value	Percent Mixed Model Estimate was same as Unaltered Data Value	Mean Absolute Difference from Unaltered Data Estimate $\lvert\hat{B}_1 - 0.0600\rvert$	
				Mixed Model Estimate	Independent Model Estimate
5	14%	66%	62%	0.00182	0.00338
9	16%	76%	70%	0.00208	0.00546
18	2%	98%	88%	0.00154	0.00932

5. CONCLUSIONS

In the examples above, we have illustrated the usefulness of RL modeling and given tests for the existence of RL censoring. The RL models can estimate regression parameters, etc. regardless of RL censoring occurring (or not). These techniques help protect an applied researcher from misanalyzing data (or

provide an additional objective assurance that independence censoring is reasonable).

The response linked censoring model of Section 2 can be generalized. Consider the RL censoring being observed at time $R_j - \varepsilon_j$ instead of at the time of the response. The ε_j can be either a constant or a random variable. Recall from Section 2, the example where an animal is attacked, resulting in the radio being destroyed (i.e., at time $R_j - \varepsilon_j$) with the animal dying later (at time P_j; the ε_j was 2 weeks).

For the use of the ε_j's being constants a likelihood very similar to (*) can be written in a straightforward fashion. Indeed, for $\varepsilon_j \to 0^+$ this generalized likelihood reduces to (*). Note that this suggests when the ε_j's are close to 0, we could use (*).

In the case of the ε_j's being random variables, it is possible to write an analogous likelihood using convolution results. The ε_j's can be assumed independent of the R_j's to make this more tractable. Essentially, the pdf of $R_j - \varepsilon_j$ is calculated and used to represent the likelihood contribution of such observations. Note that it is helpful to pick a pdf model for ε_j's such that the pdf of $R_j - \varepsilon_j$ can be a simpler, closed form expression.

Another approach is to use $R_j \gamma_j$ instead of $R_j - \varepsilon_j$ as the RL censoring time, where γ_j is a constant ($0 \leq \gamma_j \leq 1$) or a random variable. The pdf of $R_j \gamma_j$ is used similar to the pdf of $R_j - \varepsilon_j$. It should be noted that a still more general approach is to simply use a conditional pdf class given an R_j.

This paper has focused on the helpfulness of testing for RL censoring and also modeling under it. We agree with others that many important situations exist where independent censoring is the only censoring type. It is important, however, to model the difficult cases of response linked censoring that also occur. The techniques presented here allow for a better understanding of RL censoring.

ACKNOWLEDGEMENTS

Donald R. Hoover acknowledges helpful research support from a University of South Carolina Research and Productive Scholarship Grant, and the Office of Naval Research. We thank Sandra Kohta for assistance with the computer programming for this paper.

REFERENCES

Basu, A. P. and Klein, J. P. (1982). Some recent results in competing risks theory. In Survival Analysis, J. Crowley and R. A. Johnson (eds), 216-229. Hayward, California: Institute of Mathematical Statistics.

Cox, D. R. and Oakes, D. (1984). Analysis of Survival Data. London: Chapman & Hall.

David, H. A. and Moeschberger, M. L. (1978). Theory of Competing Risks. London: Griffin.

Efron, B. (1988). Logistic regression, survival analysis, and the Kaplan-Meier curve. Journal of the American Statistical Association 83, 414-425.

Gross, A. J. and Clark, V. A. (1975). Survival Distributions: Reliability Applications in the Biomedical Sciences. New York: John Wiley.

Kalbfleisch, J. and Prentice, R. (1980). The Statistical Analysis of Failure Time Data. New York: John Wiley.

Klein, J. and Moeschberger, M. (1988). Bounds on net survival probabilities for dependent competing risks. Biometrics 44, 529-538.

Lawless, J. F. (1982). Statistical Models and Methods for Lifetime Data. New York: John Wiley.

Lehman, E. (1983). Theory of Point Estimation. New York: John Wiley.

Miller, R. (1983). "What Price Kaplan-Meier?" Biometrics 39, 1077-1081.

Miller, R. (1981). Survival Analysis. New York: John Wiley.

Miller, R. and Halpern, J. (1982). Regression with Censored Data. Biometrika 63, 521-531.

Moeschberger, M. L. and Klein, J. P. (1984). Consequences of departures from independence in exponential series systems. Technometrics 26, 277-284.

Peterson, A. V. (1976). Bounds for a joint distribution function with fixed subdistribution functions: Applications to competing risks. Proceedings of the National Academy of Sciences 73, 11-13.

Pollock, K., Winterstein, S., Bunck, C., and Curtis, P. (1988a). Survival analysis in telemetry studies: the staggered entry design. To appear in Journal of Wildlife Management.

Pollock, K., Winterstein, S., and Conroy, M. (1988b). Estimation and analysis of survival distribution for radio-tagged animals. Submitted to Biometrics.

Prentice, R. L. (1973). Exponential survival with censoring and explanatory variables. Biometrika 60, 279-288.

Slud, E. and Byar, D. (1988). How dependent causes of death can make risk factors appear protective. Biometrics 44, 265-269.

Slud, E. V. and Rubinstein, L. V. (1983). Dependent competing risks and summary survival curves. Biometrika 70, 643-650.

Trent, T. T. and Rongstad, O. J. (1974). Home range and survival of cottontail rabbits in southeastern Wisconsin. Journal of Wildlife Management 38, 454-472.

Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. Proceedings of the National Academy of Sciences of the United States of America 72, 20-22.

Zippin, C. and Armitage, P. (1966). Use of concomitant variables and incomplete survival information in the estimation of an exponential survival parameter. Biometrics 22, 665-672.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 122	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Response Linked Censoring: Modeling And Estimation		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Donald R. Hoover and Frank M. Guess		8. CONTRACT OR GRANT NUMBER(s) N00014-89-J-1627
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, CA 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-267
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Code 1111		12. REPORT DATE September 14, 1989 13. NUMBER OF PAGES 18
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Office of Naval Research Statistics & Probability Program Code 1111		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Survival; Dependent Censoring; Independent Censoring; Response Linked Censoring; Covariates; Data Analysis.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

PLEASE SEE FOLLOWING PAGE.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

TECHNICAL REPORT NO. 422

20. ABSTRACT

In time to response investigations, such as survival studies, observations are often censored. The censoring may or may not be independent of the response. A particular form of positive dependence between censoring time and death time is defined as response linked (RL) censoring. Many examples are given of response linked censoring. We present a general method for modeling and estimation to be used when both response linked and independent censoring occur. This method involves parametric modeling with (or without) covariates and handles the special cases of all response linked and all independent censoring. We illustrate in detail response linked censoring and the method with the Stanford heart transplant data. Survival data on radio tagged animals provides a second example. The effectiveness of our methods is tested with a lung cancer data set where real cancer deaths have been randomly RL censored. Several other application areas (e.g., study of reformed heroin addicts time to relapse) are mentioned.